

# Sampling in the Analysis Transform Domain

Raja Giryes<sup>a</sup>

<sup>a</sup>The Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA

---

## Abstract

Many signal and image processing applications have benefited remarkably from the fact that the underlying signals reside in a low dimensional subspace. One of the main models for such a low dimensionality is the sparsity one. Within this framework there are two main options for the sparse modeling: the synthesis and the analysis ones, where the first is considered the standard paradigm for which much more research has been dedicated. In it the signals are assumed to have a sparse representation under a given dictionary. On the other hand, in the analysis approach the sparsity is measured in the coefficients of the signal after applying a certain transformation, the analysis dictionary, on it. Though several algorithms with some theory have been developed for this framework, they are outnumbered by the ones proposed for the synthesis methodology.

Given that the analysis dictionary is either a frame or the two dimensional finite difference operator, we propose a new sampling scheme for signals from the analysis model that allows recovering them from their samples using any existing algorithm from the synthesis model. The advantage of this new sampling strategy is that it makes the existing synthesis methods with their theory also available for signals from the analysis framework.

**Keywords:** Sparse representations, Compressed sensing, Synthesis, Analysis, Transform Domain.

**2010 MSC:** 94A20, 94A12, 62H12

---

## 1. Introduction

The idea that signals reside in a union of low dimensional subspaces has been used extensively in the recent decade in many fields and applications [1]. One of the main problems that has benefited remarkably from this theory is the one of compressed sensing. In this problem we want to recover an unknown signal  $\mathbf{x} \in \mathbb{R}^d$  from a small number of noisy linear measurements:

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{e}, \quad (1)$$

where  $\mathbf{M} \in \mathbb{R}^{m \times d}$  is the measurements matrix,  $\mathbf{e} \in \mathbb{R}^m$  is an additive noise and  $\mathbf{y} \in \mathbb{R}^m$  is the noisy measurement.

If the signal  $\mathbf{x}$  can be any signal then we are in a hopeless situation in the task of recovering it from  $\mathbf{y}$ . However, if we restrict it to a low-dimensional manifold that does not intersect with the null space of  $\mathbf{M}$  at any point except the origin then we are more likely to be able to recover  $\mathbf{x}$  from  $\mathbf{y}$  by looking for the signal at this manifold, which is closest to  $\mathbf{y}$  after multiplying it by  $\mathbf{M}$ .

An example for such a low dimensional manifold is the one of  $k$ -sparse signals under a given dictionary  $\mathbf{D} \in \mathbb{R}^{d \times n}$ . In this case our signal  $\mathbf{x}$  satisfies

$$\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}, \|\boldsymbol{\alpha}\|_0 \leq k, \quad (2)$$

where  $\|\boldsymbol{\alpha}\|_0$  is the  $\ell_0$ -pseudo norm that counts the number of non-zero entries in a vector. In this case we may recover  $\mathbf{x}$  from  $\mathbf{y}$  by minimizing the following problem,

$$\hat{\boldsymbol{\alpha}}_{S-\ell_0} = \underset{\tilde{\boldsymbol{\alpha}}}{\operatorname{argmin}} \|\tilde{\boldsymbol{\alpha}}\|_0 \quad s.t. \quad \|\mathbf{y} - \mathbf{M}\mathbf{D}\tilde{\boldsymbol{\alpha}}\|_2 \leq \lambda_{\mathbf{e}}. \quad (3)$$

---

Email address: raja.giryes@duke.edu. (Raja Giryes)

where  $\lambda_e$  is an upper bound for  $\|\mathbf{e}\|_2$  if the noise is bounded and adversarial, or a scalar dependent on the noise distribution [2]. As this problem is NP-hard [3] many approximation methods have been proposed for it [4, 5], such as orthogonal matching pursuit (OMP) [6] and the  $\ell_1$ -relaxation strategy that replaces the  $\ell_0$ -pseudo norm with the  $\ell_1$ -norm in (3) [7].

One of the main theoretical questions being asked with regard to these algorithms is what are the requirements on  $\mathbf{M}$ ,  $\mathbf{D}$ ,  $m$  and  $k$  such that the representation,  $\alpha$ , of  $\mathbf{x}$  may be stably recovered from  $\mathbf{y}$  using these techniques, i.e., their recovery  $\hat{\alpha}$  will satisfy

$$\|\hat{\alpha} - \alpha\|_2 \leq C \|\mathbf{e}\|_2, \quad (4)$$

where  $C$  is a certain constant (different for each algorithm).

Two main tools have been used to answer this question. The first is the coherence of  $\mathbf{MD}$  [8], which is the maximal (normalized) inner product between the columns of  $\mathbf{MD}$ . It has been shown that if the matrix  $\mathbf{MD}$  is incoherent (has a small coherence) then it is possible to get a stable recovery using OMP and the  $\ell_1$ -relaxation. The problem with the coherence based recovery conditions is that they limit the number of measurements  $m$  to be of the order of  $k^2$ , while  $m = 2k$  is enough to guarantee uniqueness for (1) in the noiseless case and  $m = O(k \log(n))$  is enough for stability in the noisy one.

The second property of  $\mathbf{MD}$  used to derive reconstruction performance guarantees is the restricted isometry property (RIP). This property provides us with a bound on the minimal and maximal eigenvalues of every sub-matrix consisting of any  $k$ -columns from a given matrix. Formally,

**Definition 1.1 (RIP [9]).** A matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  has the RIP with a constant  $\delta_k$ , if  $\delta_k$  is the smallest constant that satisfies

$$(1 - \delta_k) \|\tilde{\alpha}\|_2^2 \leq \|\mathbf{A}\tilde{\alpha}\|_2^2 \leq (1 + \delta_k) \|\tilde{\alpha}\|_2^2, \quad (5)$$

whenever  $\tilde{\alpha} \in \mathbb{R}^n$  is  $k$ -sparse.

It has been shown for many approximation algorithms that they get stable recovery in the form of (4), if  $\mathbf{MD}$  has the RIP with a constant  $\delta_{ak} < \delta_{ref}$ , where  $a$  and  $\delta_{ref}$  are two constants dependent on the algorithm in question [9, 10, 11, 12, 13, 14]. The true force behind these RIP conditions is that it has been shown that many matrices (typically random subgaussian matrices) satisfy this bound given that  $m = O(k \log(n))$  [9, 15, 16]. Notice that the main significance of this result is that it shows that it is possible to recover a signal from a number of measurements proportional to its manifold dimension  $k$ .

An alternative model for low dimensional signals that relies on sparsity is the analysis framework [17, 18]. In this paradigm, we look at the behavior of the signal after applying a certain operator  $\mathbf{\Omega} \in \mathbb{R}^{n \times d}$  on it, assuming that  $\mathbf{\Omega}\mathbf{x}$  has  $\ell$  zeros. The number of zeros,  $\ell$ , is termed the cosparsity of the signal  $\mathbf{x}$  [18]. With this prior at hand, we may recover  $\mathbf{x}$  from (1) by solving

$$\hat{\mathbf{x}}_{A-\ell_0} = \underset{\tilde{\mathbf{x}}}{\operatorname{argmin}} \|\mathbf{\Omega}\tilde{\mathbf{x}}\|_0 \quad s.t. \quad \|\mathbf{y} - \mathbf{M}\tilde{\mathbf{x}}\|_2 \leq \lambda_e, \quad (6)$$

where here also  $\lambda_e$  depends on the noise properties.

Note that as we minimize the number of non-zeros in  $\mathbf{\Omega}\mathbf{x}$  in (6), the number of zeros is the one that defines the manifold dimension in which  $\mathbf{x}$  resides. Each zero in  $\mathbf{\Omega}\mathbf{x}$  corresponds to a row in  $\mathbf{\Omega}$  to which  $\mathbf{x}$  is orthogonal. Denoting by  $T$  the support of  $\mathbf{\Omega}\mathbf{x}$  and  $T^c$  it complimentary, we may say that  $\mathbf{x}$  resides in a subspace of dimension  $d - \operatorname{rank}(\mathbf{\Omega}_{T^c})$ . Therefore if  $\mathbf{\Omega}\mathbf{x}$  has  $\ell = n - k$  zeros, where  $k$  is the number of non-zeros in it, and  $\mathbf{\Omega}$  is in general position, i.e., every  $d$  rows in it are independent, then the manifold dimension is  $d - \ell$ .

In the noiseless case ( $\mathbf{e} = 0$ ), the requirement  $m = 2(d - \operatorname{rank}(\mathbf{\Omega}_{T^c}))$  is enough to guarantee uniqueness in the solution of (6) (and therefore the recovery of  $\mathbf{x}$ ) under very mild assumptions on the relation between  $\mathbf{\Omega}$  and  $\mathbf{M}$  [18]. However, in the noisy case having a number of samples at the order of the manifold dimension, i.e.,  $m = O(d - \operatorname{rank}(\mathbf{\Omega}_{T^c}))$  is not enough to guarantee stability even by solving (6) [19]. Therefore, it is not surprising that the recovery conditions for algorithms that approximate (6) require  $m = O(k \log(n))$  [20, 21, 22, 23, 24, 25, 26], where  $\mathbf{\Omega}$  is assumed to be either a frame [20, 21, 25, 26], the 2D-DIF operator [23, 24, 27, 28] or an operator that generates a manifold with a tractable projection onto it [22].

Though the number of measurements in synthesis and analysis are similar there are two major differences between the two: (i) In synthesis the number of measurements are proportional to the manifold dimension, while in analysis this is not necessarily the case as  $k = n - \ell$  might be remarkably larger than  $d - \text{rank}(\mathbf{\Omega}_T^C)$  (See [22] for more details); (ii) In synthesis the dictionary  $\mathbf{D}$  must be incoherent as otherwise the RIP condition will no longer hold [29], while in the analysis case there is no such restriction on the analysis dictionary  $\mathbf{\Omega}$  but only on  $\mathbf{M}$ .

An interesting relation between analysis and synthesis, which is depicted in [17], is that if  $\mathbf{\Omega}$  is a frame and  $\mathbf{\Omega}\mathbf{x}$  is  $k$ -sparse then  $\mathbf{x}$  has a  $k$ -sparse representation under  $\mathbf{D} = \mathbf{\Omega}^\dagger$  (the pseudo-inverse of  $\mathbf{\Omega}$ ), i.e.,  $\mathbf{x} = \mathbf{D}\mathbf{\Omega}\mathbf{x}$ . Therefore, if  $k$  is small enough then relying on the uniqueness of the sparse representation [30], we can recover  $\mathbf{x}$  by minimizing (3). The problem we encounter in this case is that unless  $\mathbf{\Omega}$  is an incoherent matrix (its rows are incoherent) and therefore  $\mathbf{D}$  is incoherent, none of the existing synthesis approximation algorithms is guaranteed to provide us with a good estimate<sup>1</sup>.

### 1.1. Our Contribution

In this work we provide a new sampling strategy that allows recovering signals from the analysis model using any existing synthesis algorithm, given that the analysis dictionary is either a frame or the 2D-DIF operator. Our scheme is general and can be easily extended to other types of analysis dictionaries. Instead of sampling the signal itself, we sample the signal in the analysis transform domain and then perform the recovery in this domain. From the proxy in the transform domain we get a reconstruction of our original signal. The idea to recover an analysis signal in the transform domain is not a new idea and was used before [26, 37, 38, 39]. However, the uniqueness in our approach compared to previous works is that (i) we sample with one matrix and then use another one for recovery; and (ii) we make use of existing synthesis algorithms as a black box without changing them for recovering the transform domain coefficients of the signal. Our sampling and recovery strategy is presented in Section 2 for the case that the analysis dictionary is a general frame or the 2D-DIF operator. In Section 3 we provide a simple demonstration of the usage of our scheme and in Section 4 we conclude the paper.

## 2. Sampling in the Transform Domain

Before we turn to present our scheme let us recall the problem we aim at solving in the analysis case:

**Definition 2.1 (Problem  $\mathcal{P}$ ).** Consider a measurement vector  $\mathbf{y} \in \mathbb{R}^m$  such that  $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{e}$  where  $\mathbf{\Omega}\mathbf{x} \in \mathbb{R}^n$  is either  $k$ -sparse for a given and fixed analysis operator  $\mathbf{\Omega} \in \mathbb{R}^{n \times d}$  or almost  $k$ -sparse, i.e.  $\mathbf{\Omega}\mathbf{x}$  has  $k = n - \ell$  leading elements. The non-zero locations of the  $k$  leading elements is denoted by  $T$ .  $\mathbf{M} \in \mathbb{R}^{m \times d}$  is a degradation operator and  $\mathbf{e} \in \mathbb{R}^m$  is an additive noise. Our task is to recover  $\mathbf{x}$  from  $\mathbf{y}$ . The recovery result is denoted by  $\hat{\mathbf{x}}$ .

### 2.1. Guarantees for Frames

Let  $\mathbf{A} \in \mathbb{R}^{m \times n}$  be a given matrix and  $\mathcal{A}(\mathbf{y}) = \mathcal{A}(\mathbf{y}|\mathbf{A}, k)$  be an algorithm that receives a signal  $\mathbf{y}$  such that  $\mathbf{y} = \mathbf{A}\boldsymbol{\alpha} + \mathbf{e}$ , where  $\boldsymbol{\alpha} \in \mathbb{R}^n$  is either  $k$ -sparse or almost  $k$ -sparse, such that either one of the following (or the two of them) holds: (i) for the case that  $\mathbf{e}$  is an adversarial noise with a bounded energy it is guaranteed that

$$\|\boldsymbol{\alpha} - \mathcal{A}(\mathbf{y})\|_2^2 \leq C_1 \|\mathbf{e}\|_2^2 + C_2 \left( \|\boldsymbol{\alpha} - [\boldsymbol{\alpha}]_k\|_2^2 + \frac{1}{k} \|\boldsymbol{\alpha} - [\boldsymbol{\alpha}]_k\|_1^2 \right), \quad (7)$$

where  $[\boldsymbol{\alpha}]_k$  is the best  $k$ -term approximation of  $\boldsymbol{\alpha}$ , and  $C_1$  and  $C_2$  are two constants depending on  $\mathbf{A}$  and the algorithms<sup>2</sup> (See [9, 10, 11, 12, 13, 14]); or (ii) for the case that  $\mathbf{e}$  is a zero-mean white Gaussian noise with variance  $\sigma^2$ , it is guaranteed that with a high probability,

$$\|\boldsymbol{\alpha} - \mathcal{A}(\mathbf{y})\|_2^2 \leq C_3 k \sigma^2 \log(n) + C_4 \left( \|\boldsymbol{\alpha} - [\boldsymbol{\alpha}]_k\|_2^2 + \frac{1}{k} \|\boldsymbol{\alpha} - [\boldsymbol{\alpha}]_k\|_1^2 \right), \quad (8)$$

where  $C_3$  and  $C_4$  are two constants depending on  $\mathbf{A}$  and the algorithms (See [40, 41, 42, 43]).

Assuming that  $\mathbf{\Omega}$  in Problem  $\mathcal{P}$  is a frame, we propose the following sampling and reconstruction strategy:

<sup>1</sup>Some recent works have addressed the case of coherent dictionaries in the synthesis case [31, 32, 33, 34, 35, 36]. However, they are very limited to specific cases and do not apply to general types of dictionaries such as frames.

<sup>2</sup>Note that (7) is a generalization of the bound in (4) for the case that  $\boldsymbol{\alpha}$  is a non-exact  $k$ -sparse vector.

- Set the sensing matrix to be  $\mathbf{M} = \mathbf{A}\mathbf{\Omega}$ . In this case we have  $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{e} = \mathbf{A}\mathbf{\Omega}\mathbf{x} + \mathbf{e}$  and therefore we can apply algorithm  $\mathcal{A}$  to recover  $\mathbf{\Omega}\mathbf{x}$  as it is a  $k$ -sparse (or approximately so) vector.
- Compute an estimate for  $\mathbf{\Omega}\mathbf{x}$ :  $\hat{\mathbf{a}} = \mathcal{A}(\mathbf{y})$ .
- Use the frame's Moore-Penrose pseudo-inverse to recover  $\mathbf{x}$ :  $\hat{\mathbf{x}} = \mathbf{\Omega}^\dagger \hat{\mathbf{a}}$ .

This algorithm is summarized also in Algorithm 1. Remark that we sample in the transform domain of  $\mathbf{\Omega}$ , as we sample with  $\mathbf{M} = \mathbf{A}\mathbf{\Omega}$ , and then recover only with  $\mathbf{A}$  the transform coefficients of  $\mathbf{x}$ , i.e.  $\mathbf{\Omega}\mathbf{x}$ . Note also that in the final step, where we calculate  $\hat{\mathbf{x}} = \mathbf{\Omega}^\dagger \hat{\mathbf{w}}$ , we may replace  $\mathbf{\Omega}^\dagger$  with any dictionary that satisfies  $\mathbf{D}\mathbf{\Omega} = \mathbf{I}$ .

---

**Algorithm 1** Signal Recovery from Samples of Frames in the Transform Domain

---

**Require:**  $k, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{\Omega} \in \mathbb{R}^{n \times d}, \mathbf{y}, \mathcal{A}$ , where  $\mathbf{y} = \mathbf{A}\mathbf{\Omega}\mathbf{x} + \mathbf{e}$ ,  $\mathbf{\Omega}\mathbf{x}$  is a  $k$ -sparse vector or approximately so,  $\mathbf{e}$  is an additive noise, and  $\mathcal{A}(\cdot) = \mathcal{A}(\cdot|\mathbf{A}, k)$  is a synthesis recovery program for  $k$ -sparse signals under the matrix  $\mathbf{A}$ .

**Ensure:**  $\hat{\mathbf{x}}$ : Approximation of  $\mathbf{x}$ .

Get a transform domain proxy for  $\mathbf{\Omega}\mathbf{x}$ :  $\hat{\mathbf{w}} = \mathcal{A}(\mathbf{y}|\mathbf{A}, k)$

Signal recovery:  $\hat{\mathbf{x}} = \mathbf{\Omega}^\dagger \hat{\mathbf{w}}$ , generating a signal estimate using the transform domain proxy.

---

The following theorem provides guarantees for signal recovery using the above scheme given that the synthesis reconstruction program used in it  $\mathcal{A}$  satisfies either (7) or (8), or both of them.

**Theorem 2.2 (Signal recovery from samples of frames in the transform domain).** *Consider the problem  $\mathcal{P}$  such that  $\mathbf{M} = \mathbf{A}\mathbf{\Omega}$  and  $\mathbf{\Omega}$  is a frame with a lower frame bound  $A$ . Let  $\hat{\mathbf{x}}$  be the output of Algorithm 1 with the synthesis program  $\mathcal{A}(\cdot|\mathbf{A}, k)$ . If  $\mathbf{e}$  is a bounded additive adversarial noise and (7) holds for  $\mathcal{A}(\cdot|\mathbf{A}, k)$  then*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq \frac{C_1}{A^2} \|\mathbf{e}\|_2^2 + \frac{C_2}{A^2} \left( \|\mathbf{\Omega}_{T^c} \mathbf{x}\|_2^2 + \frac{1}{k} \|\mathbf{\Omega}_{T^c} \mathbf{x}\|_1^2 \right), \quad (9)$$

*implying a stable recovery. If  $\mathbf{e}$  is a zero-mean white Gaussian noise with variance  $\sigma^2$  and (8) holds for  $\mathcal{A}(\cdot|\mathbf{A}, k)$  then with a high probability,<sup>3</sup>*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq \frac{C_3}{A^2} k \log n \sigma^2 + \frac{C_4}{A^2} \left( \|\mathbf{\Omega}_{T^c} \mathbf{x}\|_2^2 + \frac{1}{k} \|\mathbf{\Omega}_{T^c} \mathbf{x}\|_1^2 \right), \quad (10)$$

*implying a denoising effect. The constants  $C_1, C_2, C_3, C_4$  are the same as in (7) and (8).*

*Proof:* We prove only the bound in (9). The proof for (10) is very similar and omitted. Assume that (7) holds. Then since  $\mathbf{y} = \mathbf{A}\mathbf{\Omega}\mathbf{x} + \mathbf{e}$ , we have that

$$\|\mathbf{\Omega}\mathbf{x} - \hat{\mathbf{w}}\|_2^2 = \|\mathbf{\Omega}\mathbf{x} - \mathcal{A}(\mathbf{y})\|_2^2 \leq C_1 \|\mathbf{e}\|_2^2 + C_2 \left( \|\mathbf{\Omega}\mathbf{x} - [\mathbf{\Omega}\mathbf{x}]_k\|_2^2 + \frac{1}{k} \|\mathbf{\Omega}\mathbf{x} - [\mathbf{\Omega}\mathbf{x}]_k\|_1^2 \right). \quad (11)$$

We get (9) by using the facts that (i)  $[\mathbf{\Omega}\mathbf{x}]_k = \mathbf{\Omega}_T \mathbf{x}$  and therefore  $\mathbf{\Omega}\mathbf{x} - [\mathbf{\Omega}\mathbf{x}]_k = \mathbf{\Omega}_{T^c} \mathbf{x}$ ; (ii)  $\mathbf{\Omega}$  is a frame with a lower frame bound  $A$  and therefore  $\|\mathbf{\Omega}^\dagger\|_2 \leq \frac{1}{A}$ ; and (iii)  $\mathbf{x} = \mathbf{\Omega}^\dagger \mathbf{\Omega}\mathbf{x}$  and thus  $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 = \|\mathbf{\Omega}^\dagger (\mathbf{\Omega}\mathbf{x} - \hat{\mathbf{w}})\|_2$ .  $\square$

This theorem provides the same guarantees derived for analysis algorithms, which were designed especially for the analysis framework, using already existing methods from the synthesis model. The wide use of the latter and the large variety of programs available for it allow recovering a signal from a small number of measurements with more ease, using our new sampling scheme. In addition, we may say that the above theorem demonstrates that our new sampling scheme allows transferring almost any existing result from the synthesis framework to the analysis one. One example is the ability to set  $\mathbf{A}$  to be an expander graph. In this case, it is possible to recover the signal  $\mathbf{x}$  using only  $k$  steps [44]. To the best of our knowledge, such an efficient strategy does not exist for the analysis framework.

---

<sup>3</sup>Remark that it is also possible to provide guarantees for the expectation of the error, given a variant of (8) that bounds the expectation of the error like in [40, 43]

## 2.2. Guarantees for the 2D-DIF Operator

Having a guarantee for frames we turn to provide a guarantee for 2D-DIF, the two-dimensional finite difference operator. For convenience we assume that  $\mathbf{x}$  is an image (column stacked) of size  $N \times N = d$  ( $N = \sqrt{d}$ ). Notice that unlike frames, for the 2D-DIF operator a small distance in the transform domain does not imply a small distance in the signal domain. For example, the distance between two constant images is zero in the transform domain of the 2D-DIF operator. However, it can be arbitrarily large as we want depending on the constant value we assign to each image. Therefore, it is impossible to recover a signal by just using the scheme we have in Algorithm 1. Note that the problem lies in the last stage of the algorithm as we do not have enough information to get back stably from the transform domain to the signal domain. Note that also if we will add rows to  $\mathbf{\Omega}_{2D-DIF}$  and then apply a pseudo inverse, we will not have a stable recovery in the signal domain given the recovery in the transform domain (See [23, 24] for more details).

Therefore we utilize the tools used in [23] that studies the performance of the 2D-DIF operator with the analysis  $\ell_1$ -minimization, which is known also as the anisotropic total variation (TV). Two key steps are used in that work for developing the result for TV:

- The construction of the measurements:

$$\mathbf{y} = \begin{pmatrix} \mathbf{M}_1 \mathbf{x}_{nfr} \\ \mathbf{M}_1 \mathbf{x}_{nlr} \\ \mathbf{M}_2 \mathbf{x}_{nfc} \\ \mathbf{M}_2 \mathbf{x}_{nlc} \\ \mathbf{B} \mathbf{x} \end{pmatrix} + \mathbf{e}, \quad (12)$$

where  $\mathbf{x}_{nfr}$ ,  $\mathbf{x}_{nlr}$ ,  $\mathbf{x}_{nfc}$  and  $\mathbf{x}_{nlc}$  are versions of  $\mathbf{x}$  with no first row, last row, first column or last column respectively. In addition,  $\mathbf{M}_1, \mathbf{M}_2 \in \mathbb{R}^{m_1 \times N(N-1)}$  are assumed to satisfy the RIP with  $\delta_{5k} < \frac{1}{3}$  and  $\mathbf{B}\mathbf{H}^{-1}$  is assumed to satisfy the RIP with  $\delta_{2k} < 1$ , where  $\mathbf{H}$  is the bivariate Haar transform and  $\mathbf{B} \in \mathbb{R}^{m_2 \times d}$ .

- The usage of the relationship between  $\mathbf{\Omega}_{2D-DIF}$  and  $\mathbf{H}$ : For any vector  $\mathbf{v}$ , if  $\|\mathbf{\Omega}_{2D-DIF} \mathbf{v}\|_0 \leq k$  then  $\|\mathbf{H}\mathbf{v}\|_0 \leq k \log(d)$ .

The first two measurement matrices  $\mathbf{M}_1$  and  $\mathbf{M}_2$  provide information about the derivatives of  $\mathbf{x}$  and lead to a stable recovery of  $\mathbf{\Omega}_{2D-DIF} \mathbf{x}$ , the discrete gradient vector of  $\mathbf{x}$ . As we have mentioned before  $\mathbf{\Omega}_{2D-DIF}$  is non-invertible. Therefore, the reconstruction of the derivatives is not enough for recovering the signal. For this purpose the third matrix  $\mathbf{B}$  is used to guarantee stable recovery also in the signal domain. This is achieved using the following theorem:

**Theorem 2.3 (Strong Sobolev inequality. Theorem 8 in [23]).** *Let  $N$  be a power of 2 and  $\mathbf{B}$  be a linear map which, composed with the inverse bivariate Haar transform  $\mathbf{B}\mathbf{H}^{-1} \in \mathbb{R}^{m_2 \times d}$ , has the RIP with a constant  $\delta_{2k} < 1$ . Suppose that for  $\mathbf{z} \in \mathbb{R}^d$  we have  $\|\mathbf{B}\mathbf{z}\|_2 \leq \epsilon$ . Then*

$$\|\mathbf{z}\|_2 \leq \frac{2C_{\mathbf{H}}}{1 - \delta_{2k}} \frac{1}{\sqrt{k}} \log(d/k) \|\mathbf{\Omega}_{2D-DIF} \mathbf{z}\|_1 + \frac{1}{1 - \delta_{2k}} \epsilon, \quad (13)$$

where  $C_{\mathbf{H}} = 36(480\sqrt{5} + 168\sqrt{3})$ .

We utilize the above theorem for extending our sampling technique for the 2D-DIF operator. By observing again the samples generated by  $\mathbf{M}_1$  and  $\mathbf{M}_2$ , and denoting by  $\mathbf{\Omega}_v$  and  $\mathbf{\Omega}_h$  the vertical and horizontal difference of  $\mathbf{\Omega}_{2D-DIF}$  respectively, we can write  $\mathbf{M}_1 \mathbf{x}_{nfr} - \mathbf{M}_1 \mathbf{x}_{nlr} = \mathbf{M}_1 \mathbf{\Omega}_v \mathbf{x}$  and  $\mathbf{M}_1 \mathbf{x}_{nfc} - \mathbf{M}_1 \mathbf{x}_{nlc} = \mathbf{M}_2 \mathbf{\Omega}_h \mathbf{x}$ . Alternatively, we can rewrite it as

$$\begin{pmatrix} \mathbf{M}_1 & 0 \\ 0 & \mathbf{M}_2 \end{pmatrix} \mathbf{\Omega}_{2D-DIF} \mathbf{x}, \quad (14)$$

and we end up with having samples from the derivatives domain. Notice that we do not have to restrict ourselves to a block diagonal matrix composed of two linear maps for sampling each derivative direction. We can use any sampling

operator that has recovery guarantees in the synthesis framework for reconstructing the coefficients in the transform domain. We denote this reconstruction by  $\hat{\mathbf{w}}$ .

In order to recover the signal from its proxy  $\hat{\mathbf{w}}$ , we take more measurements of the original signal  $\mathbf{x}$ . These are taken using a matrix  $\mathbf{B}$  for which  $\mathbf{B}\mathbf{H}^{-1}$  (its composition with the inverse bivariate Haar transform) has the RIP with a constant  $\delta_{2k} < 1$ . Given these measurements,  $\mathbf{y}_2 = \mathbf{B}\mathbf{x} + \mathbf{e}_2$ , we get a recovery of the signal by solving

$$\hat{\mathbf{x}}_{2D-DIF} = \underset{\tilde{\mathbf{x}}}{\operatorname{argmin}} \|\mathbf{\Omega}\tilde{\mathbf{x}} - \hat{\mathbf{w}}_{2D-DIF}\|_1 \quad s.t. \quad \|\mathbf{B}\tilde{\mathbf{x}} - \mathbf{y}_2\|_2 \leq \|\mathbf{e}_2\|_2. \quad (15)$$

---

**Algorithm 2** Signal Recovery from Samples of 2D-DIF in the Transform Domain

---

**Require:**  $k, \mathbf{A} \in \mathbb{R}^{m_1 \times n}, \mathbf{B} \in \mathbb{R}^{m_2 \times d}, \mathbf{\Omega}_{2D-DIF}, \mathbf{y}, \mathcal{A}$ , where  $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$  such that  $\mathbf{y}_1 = \mathbf{A}\mathbf{\Omega}_{2D-DIF}\mathbf{x} + \mathbf{e}_1$  and  $\mathbf{y}_2 = \mathbf{B}\mathbf{x} + \mathbf{e}_2$ ,  $\mathbf{\Omega}_{2D-DIF}\mathbf{x}$  is  $k$  sparse or approximately so,  $\mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$  is an additive noise, and  $\mathcal{A}(\cdot) = \mathcal{A}(\cdot, \mathbf{A}, k)$  is a synthesis recovery program for  $k$ -sparse representation under the matrix  $\mathbf{A}$ .  
**Ensure:**  $\hat{\mathbf{x}}$ : Approximation of  $\mathbf{x}$ .  
 Get a transform domain proxy for  $\mathbf{\Omega}\mathbf{x}$ :  $\hat{\mathbf{w}} = \mathcal{A}(\mathbf{y}_1, \mathbf{A}, k)$   
 Signal recovery: Calculate  $\hat{\mathbf{x}}$  using (15) with  $\mathbf{y}_2$  and  $\hat{\mathbf{w}}$ .

---

To sum it up, our sampling strategy for the 2D-DIF operator consists of taking two sets of measurements. The first in the transform domain,  $\mathbf{y}_1 = \mathbf{A}\mathbf{\Omega}_{2D-DIF}\mathbf{x} + \mathbf{e}_1$ , leads to reconstruction of the gradient components. The second is taken with a linear map which is well behaved if applied together with the inverse of the bivariate Haar,  $\mathbf{y}_2 = \mathbf{B}\mathbf{x} + \mathbf{e}_2$ , where its sole purpose is to convert the transform domain estimate into a signal estimate using (15). Note that the linear map we use for sampling is  $\mathbf{M} = \begin{bmatrix} \mathbf{A}\mathbf{\Omega}_{2D-DIF} \\ \mathbf{B} \end{bmatrix}$  and our measurements are of the form  $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{e}$ , where  $\mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$ . Our recovery strategy from these samples is summarized in Algorithm 2. Note that in (15) we can use  $\|\mathbf{e}\|_2$  instead of  $\|\mathbf{e}_2\|_2$  if we do not have a good bound for the latter.

For the theoretical study of Algorithm 2 we make a different assumption on the used synthesis program  $\mathcal{A}$ . Instead of the bounds in (7) and (8) we assume that the following holds:

$$\|\alpha - \mathcal{A}(\mathbf{y})\|_1 \leq C_5 \sqrt{k} \|\mathbf{e}\|_2 + C_6 \|\alpha - [\alpha]_k\|_1. \quad (16)$$

Such a bound holds for the synthesis  $\ell_1$ -minimization with RIP matrices [23]. With this assumption we are ready to introduce the recovery guarantee for Algorithm 2.

**Theorem 2.4 (Stable signal recovery from samples of 2D-DIF in the transform domain).** *Consider the problem  $\mathcal{P}$  such that  $\mathbf{M} = \begin{bmatrix} \mathbf{A}\mathbf{\Omega}_{2D-DIF} \\ \mathbf{B} \end{bmatrix}$ , where  $\mathbf{A}$  has the RIP with a constant  $\delta_{ak}$  for a certain constant  $a \geq 1$ ,  $\mathbf{\Omega}_{2D-DIF}$  is the 2D-DIF operator and  $\mathbf{B}\mathbf{H}^{-1}$  has the RIP with a constant  $\delta_{2k} < 1$ . Let  $\hat{\mathbf{x}}$  be the output of Algorithm 2 with the synthesis program  $\mathcal{A}(\cdot|\mathbf{A}, k)$ . If  $\mathbf{e}$  is a bounded additive adversarial noise and (16) holds for  $\mathcal{A}(\cdot|\mathbf{A}, k)$  then*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \log(d/k) \left( C_7 \|\mathbf{e}\|_2 + \frac{C_8}{\sqrt{k}} \|\mathbf{\Omega}_{T^c} \mathbf{x}\|_1 \right), \quad (17)$$

implying a stable recovery, where  $C_7$  and  $C_8$  are functions of  $C_{\mathbf{H}}$  and  $\delta_{2k}$ .

*Proof:* Since  $\hat{\mathbf{x}}$  is a minimizer of (15) we have that

$$\|\mathbf{B}\hat{\mathbf{x}} - \mathbf{y}_2\|_2 \leq \|\mathbf{e}_2\|_2. \quad (18)$$

Since  $\mathbf{y}_2 = \mathbf{B}\mathbf{x} + \mathbf{e}_2$  we have from the triangle inequality that

$$\|\mathbf{B}(\hat{\mathbf{x}} - \mathbf{x})\|_2 \leq 2 \|\mathbf{e}_2\|_2. \quad (19)$$

Therefore, setting  $\mathbf{z} = \hat{\mathbf{x}} - \mathbf{x}$  in Theorem 2.3 we have

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{2C_H}{1 - \delta_{2k}} \frac{1}{\sqrt{k}} \log(d/k) \|\mathbf{\Omega}_{2D-DIF}(\hat{\mathbf{x}} - \mathbf{x})\|_1 + \frac{2}{1 - \delta_{2k}} \|\mathbf{e}_2\|_2. \quad (20)$$

From the triangle inequality we have

$$\|\mathbf{\Omega}_{2D-DIF}(\hat{\mathbf{x}} - \mathbf{x})\|_1 \leq \|\mathbf{\Omega}_{2D-DIF} \hat{\mathbf{x}} - \hat{\mathbf{w}}\|_1 + \|\hat{\mathbf{w}} - \mathbf{\Omega}_{2D-DIF} \mathbf{x}\|_1. \quad (21)$$

Since  $\mathbf{x}$  is a feasible solution to (15) and  $\hat{\mathbf{x}}$  is its minimizer we have

$$\|\mathbf{\Omega}_{2D-DIF} \hat{\mathbf{x}} - \hat{\mathbf{w}}\|_1 \leq \|\mathbf{\Omega}_{2D-DIF} \mathbf{x} - \hat{\mathbf{w}}\|_1. \quad (22)$$

Plugging (22) in (21) we have

$$\|\mathbf{\Omega}_{2D-DIF}(\hat{\mathbf{x}} - \mathbf{x})\|_1 \leq 2 \|\mathbf{\Omega}_{2D-DIF} \mathbf{x} - \hat{\mathbf{w}}\|_1. \quad (23)$$

Notice that we can bound the right hand side (rhs) of (23) with (16), where  $\boldsymbol{\alpha} = \mathbf{\Omega} \mathbf{x}$  and  $\hat{\boldsymbol{\alpha}} = \hat{\mathbf{w}}$ . Therefore, by combining (23) and (16) with (20) we have

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{2C_H}{1 - \delta_{2k}} \log(d/k) \left( C_5 \|\mathbf{e}\|_2 + \frac{C_6}{\sqrt{k}} \|\mathbf{\Omega}_{T^c} \mathbf{x}\|_1 \right) + \frac{2}{1 - \delta_{2k}} \|\mathbf{e}_2\|_2. \quad (24)$$

□

**Remark 2.5.** An example of a procedure  $\mathcal{A}(\cdot | \mathbf{A}, k)$ , for which (16) holds, is the synthesis  $\ell_1$ -minimization with  $\mathbf{A}$  having the RIP with a constant  $\delta_{5k} \leq \frac{1}{3}$  [23].

### 2.3. Guarantees for a General Analysis Operator

---

#### Algorithm 3 Signal Recovery from Samples of a General Analysis Operator in the Transform Domain

---

**Require:**  $k, \mathbf{A} \in \mathbb{R}^{m_1 \times n}, \mathbf{B} \in \mathbb{R}^{m_2 \times d}, \mathbf{\Omega}, \mathbf{y}, p, \mathcal{A}$ , where  $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$  such that  $\mathbf{y}_1 = \mathbf{A} \mathbf{\Omega} \mathbf{x} + \mathbf{e}_1$  and  $\mathbf{y}_2 = \mathbf{B} \mathbf{x} + \mathbf{e}_2$ ,  $\mathbf{\Omega} \mathbf{x}$  is  $k$  sparse or approximately so,  $\mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$  is an additive noise,  $p$  is the  $\ell_p$  norm used in this algorithm, and  $\mathcal{A}(\cdot) = \mathcal{A}(\cdot, \mathbf{A}, k)$  is a synthesis recovery program for  $k$ -sparse representation under the matrix  $\mathbf{A}$ .

**Ensure:**  $\hat{\mathbf{x}}$ : Approximation of  $\mathbf{x}$ .

Get a transform domain proxy for  $\mathbf{\Omega} \mathbf{x}$ :  $\hat{\mathbf{w}} = \mathcal{A}(\mathbf{y}_1, \mathbf{A}, k)$ .

Signal recovery:

$$\hat{\mathbf{x}} = \underset{\tilde{\mathbf{x}}}{\operatorname{argmin}} \|\mathbf{\Omega} \tilde{\mathbf{x}} - \hat{\mathbf{w}}\|_p \quad s.t. \quad \|\mathbf{B} \tilde{\mathbf{x}} - \mathbf{y}_2\|_2 \leq \|\mathbf{e}_2\|_2. \quad (25)$$


---

Extending this idea further we do not restrict the sampling strategy in Algorithm 2 only to  $\mathbf{\Omega}_{2D-DIF}$ . We present this extension in Algorithm 3. It can be applied for any operator for which a stable recovery in the coefficients domain implies a stable recovery in the signal domain by some additional measurements of the signal. The following theorem, which is similar to Theorem 2.4, provides a recovery guarantee for this generalized scheme.



**Theorem 2.6 (Stable signal recovery from samples of a general analysis operator in the transform domain).** Consider the problem  $\mathcal{P}$  such that  $\mathbf{M} = \begin{bmatrix} \mathbf{A}\mathbf{\Omega} \\ \mathbf{B} \end{bmatrix}$ , where  $\mathbf{A} \in \mathbb{R}^{m_1 \times n}$  and  $\mathbf{\Omega}$  is a general analysis operator. Suppose  $\mathbf{B}$  is a matrix such that for any  $\mathbf{z} \in \mathbb{R}^d$ ,  $\|\mathbf{B}\mathbf{z}\|_2 \leq \epsilon$  implies

$$\|\mathbf{z}\|_2 \leq \beta \|\mathbf{\Omega}\mathbf{z}\|_p + \gamma\epsilon. \quad (26)$$

and that for any  $\alpha \in \mathbb{R}^n$  and  $\mathbf{y}_1 \in \mathbb{R}^{m_1}$

$$\|\alpha - \mathcal{A}(\mathbf{y})\|_p \leq \zeta \|\mathbf{e}\|_2 + \xi \|\alpha - [\alpha]_k\|_1. \quad (27)$$

holds for the synthesis program  $\mathcal{A}(\cdot|\mathbf{A}, k)$ . Let  $\hat{\mathbf{x}}$  be the output of Algorithm 3 with the program  $\mathcal{A}(\cdot|\mathbf{A}, k)$  and  $\mathbf{e}$  be a bounded additive adversarial noise. Then

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq 2\beta(\zeta \|\mathbf{e}\|_2 + \xi \|\mathbf{\Omega}_{T^c}\mathbf{x}\|_1) + 2\gamma \|\mathbf{e}_2\|_2. \quad (28)$$

*Proof:* As the proof is very similar to the one of Theorem 2.4 we present it briefly. Using the same steps that led to (20) and (23) we have

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq 2\beta \|\mathbf{\Omega}(\hat{\mathbf{x}} - \mathbf{x})\|_p + 2\gamma \|\mathbf{e}_2\|_2. \quad (29)$$

and

$$\|\mathbf{\Omega}(\hat{\mathbf{x}} - \mathbf{x})\|_p \leq 2 \|\mathbf{\Omega}\mathbf{x} - \hat{\mathbf{w}}\|_p. \quad (30)$$

Plugging (27) in (30), with  $\alpha = \mathbf{\Omega}\mathbf{x}$ , and then combining the result with (29) lead to (28).  $\square$

Notice that the result in Theorem 2.4 is a special case of the above theorem. We present two other special cases in the following two corollaries. The first is a generalization of Theorem 2.4 for  $L$ -dimensional signals and the LD-DIF operator, the  $L$  dimensional finite difference analysis dictionary.

**Corollary 2.7 (Stable signal recovery from samples of LD-DIF in the transform domain).** Consider the problem  $\mathcal{P}$  such that  $\mathbf{M} = \begin{bmatrix} \mathbf{A}\mathbf{\Omega}_{LD-DIF} \\ \mathbf{B} \end{bmatrix}$ , where  $\mathbf{\Omega}_{LD-DIF}$  is the LD-DIF operator;  $\mathbf{B}\mathbf{H}^{-1}$  has the RIP with a constant  $\delta_{2k} < 1$ , and  $\mathbf{H}$  is the  $L$ -dimensional Haar wavelet transform. Let  $\hat{\mathbf{x}}$  be the output of Algorithm 3 with the synthesis program  $\mathcal{A}(\cdot|\mathbf{A}, k)$  and  $p = 1$ . If  $\mathbf{e}$  is a bounded additive adversarial noise and (16) holds for  $\mathcal{A}(\cdot|\mathbf{A}, k)$  then

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \log(d) \left( C_9 \|\mathbf{e}\|_2 + \frac{C_{10}}{\sqrt{k}} \|\mathbf{\Omega}_{T^c}\mathbf{x}\|_1 \right), \quad (31)$$

implying a stable recovery, where  $C_9$  and  $C_{10}$  are certain constants.

The proof follows from a generalized version of Theorem 2.3 for the  $L$ -dimensional case (Theorem 6 in [24]) that provides (26) with  $\gamma = 1$  and  $\beta = \log(d) \frac{C}{\sqrt{k}}$ , where  $C$  is a certain constant.

**Remark 2.8.** Notice that one may further generalize Theorem 2.6 to deal also with block sparsity [45, 46, 47, 48], i.e., the case that  $\mathbf{\Omega}_1\mathbf{x}, \mathbf{\Omega}_2\mathbf{x}, \dots, \mathbf{\Omega}_L\mathbf{x}$  are jointly sparse, where  $\mathbf{\Omega} = [\mathbf{\Omega}_1^T, \mathbf{\Omega}_2^T, \dots, \mathbf{\Omega}_L^T]^T$ . In this case, the  $\ell_1$ -norm applied on vectors in  $\mathbb{R}^n$  in Algorithm 3, Theorem 2.6 and (16) needs to be replaced with the mixed  $\ell_{1,2}$ -norm<sup>4</sup> applied on matrices in  $\mathbb{R}^{\frac{n}{L} \times L}$ . An example for such a case is the  $L$ -dimensional isotropic total variation, where  $\mathbf{\Omega}_i$  is the derivative in the  $i$ -th dimension (if  $L = 2$  then  $\mathbf{\Omega}_1$  and  $\mathbf{\Omega}_2$  are the horizontal and vertical derivatives respectively). Note that it can be shown that the  $\ell_{1,2}$ -minimization algorithm satisfies a version of (16) with the  $\ell_{1,2}$ -norm. In addition, Theorem 6 in [24] provides a bound in the form of (26) with the  $\ell_{1,2}$ -norm instead of the  $\ell_1$ -norm. Therefore, it is possible to derive a theorem similar to Corollary 2.7 equivalent to the theorems for the isotropic TV in [24]. We leave the details to the interested reader.

<sup>4</sup>Applying an  $\ell_2$ -norm on the rows followed by an  $\ell_1$ -norm on the resulted vector.



The second corollary considers operators that can be viewed as part of a frame.

**Corollary 2.9 (Stable signal recovery from samples of a partial frame).** *Consider the problem  $\mathcal{P}$  such that  $\mathbf{M} = \begin{bmatrix} \mathbf{A}\mathbf{\Omega} \\ \mathbf{B} \end{bmatrix}$ , where  $\mathbf{\Omega}$  is a matrix for which there exists  $\tilde{\mathbf{\Omega}}$  such that  $\mathbf{\Omega}_F = [\mathbf{\Omega}^T, \tilde{\mathbf{\Omega}}^T]^T$  is a frame with a lower frame bound  $A$ , and  $\sigma_{\min}(\mathbf{B}\tilde{\mathbf{\Omega}}^\dagger) \geq C_{11}$  and  $\|\mathbf{B}(\mathbf{I} - \tilde{\mathbf{\Omega}}^\dagger\tilde{\mathbf{\Omega}})\| \leq C_{12}$  for constants  $C_{11}$  and  $C_{12}$  satisfying  $\frac{C_{12}}{C_{11}A} < 1$ . Let  $\hat{\mathbf{x}}$  be the output of Algorithm 3 with the synthesis program  $\mathcal{A}(\cdot|\mathbf{A}, k)$  and  $p = 2$ . If  $\mathbf{e}$  is a bounded additive adversarial noise and (27) holds for  $\mathcal{A}(\cdot|\mathbf{A}, k)$  with  $\zeta = C_{13}$  and  $\xi = \frac{C_{14}}{\sqrt{k}}$ , where  $C_{13}$  and  $C_{14}$  are certain constants, then*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \left( C_{15} \|\mathbf{e}\|_2 + \frac{C_{16}}{\sqrt{k}} \|\mathbf{\Omega}_T^c \mathbf{x}\|_1 \right), \quad (32)$$

implying a stable recovery, where  $C_{15}$  and  $C_{16}$  are constants dependent only on  $A$ ,  $C_{11}$ ,  $C_{12}$ ,  $C_{13}$  and  $C_{14}$ .

*Proof:* For the proof we just need to show that (26) holds. Using the lower frame bound followed by the triangle inequality and the fact that  $\sigma_{\min}(\mathbf{B}\tilde{\mathbf{\Omega}}^\dagger) \geq C_{11}$ , we have

$$\|\mathbf{z}\|_2 \leq \frac{1}{A} \|\mathbf{\Omega}_F \mathbf{z}\|_2 \leq \frac{1}{A} \|\mathbf{\Omega} \mathbf{z}\|_2 + \frac{1}{A} \|\tilde{\mathbf{\Omega}} \mathbf{z}\|_2 \leq \frac{1}{A} \|\mathbf{\Omega} \mathbf{z}\|_2 + \frac{1}{AC_{11}} \|\mathbf{B}\tilde{\mathbf{\Omega}}^\dagger \tilde{\mathbf{\Omega}} \mathbf{z}\|_2 \quad (33)$$

Using the triangle inequality and the fact that  $\|\mathbf{B}(\mathbf{I} - \tilde{\mathbf{\Omega}}^\dagger\tilde{\mathbf{\Omega}})\| \leq C_{12}$  we have

$$\|\mathbf{B}\tilde{\mathbf{\Omega}}^\dagger \tilde{\mathbf{\Omega}} \mathbf{z}\|_2 \leq \|\mathbf{B} \mathbf{z}\|_2 + C_{12} \|\mathbf{z}\|_2 \quad (34)$$

Plugging (34) in (33) with some simple arithmetical steps lead to

$$\left( 1 - \frac{C_{12}}{AC_{11}} \right) \|\mathbf{z}\|_2 \leq \frac{1}{A} \|\mathbf{\Omega} \mathbf{z}\|_2 + \frac{1}{AC_{11}} \|\mathbf{B} \mathbf{z}\|_2. \quad (35)$$

Notice that by the assumptions of the corollary  $1 - \frac{C_{12}}{AC_{11}} > 0$ . This equation provides the constants in (26), completing the proof.  $\square$

**Remark 2.10.** An example for a program  $\mathcal{A}(\cdot|\mathbf{A}, k)$  that satisfies the assumption of the theorem is CoSaMP [10].

**Remark 2.11.** An example for a matrix  $\mathbf{B}$  that satisfies the assumptions of the theorem is  $\mathbf{B} = \tilde{\mathbf{\Omega}}$ . In this case  $C_{11} = 1$  and  $C_{12} = 0$ .

Another family of analysis operators that might be of interest is the one of convolutional operators [49]. In this case the condition number of  $\mathbf{\Omega}$  is usually very large and the sampling strategy used with Algorithm 3 is needed, as we cannot sample directly from the transform domain like in the case of frames. We leave the exploration of this case to a future research.

### 3. Epilogue - Do We Still Need Analysis Algorithms?

Following the fact that our proposed recovery guarantees are similar to the ones achieved for the existing analysis algorithms and that sampling in the manifold dimension of analysis signals lead to unstable recovery [19], one may ask whether there is a need at all for reconstruction strategies that rely on the analysis model. For this reason we perform several experiments to compare the empirical recovery performance of our new sampling scheme, with synthesis  $\ell_1$ -minimization, and the standard sampling scheme, with analysis  $\ell_1$ -minimization, for signals from the analysis framework. The minimizations are performed using *cvx* [50, 51].

We start with the case of signals that are sparse after applying randomly generated tight-frames. We set  $\mathbf{\Omega} \in \mathbb{R}^{144 \times 120}$ , where the signal dimension is  $d = 200$ , and  $k = 144 - 110$  (setting the signal intrinsic dimension to

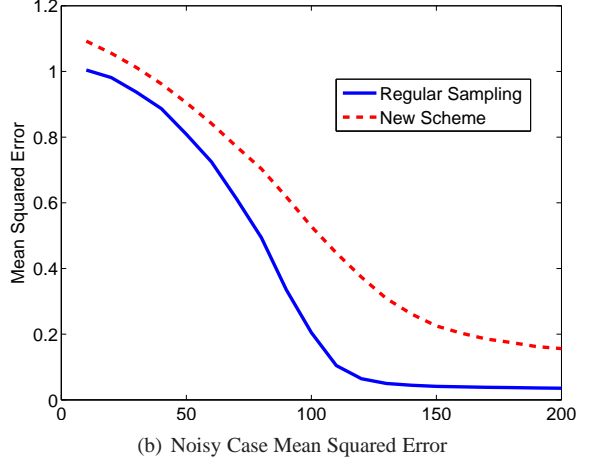
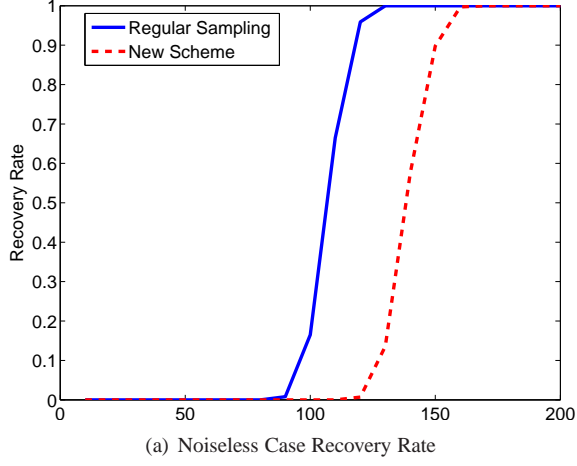


Figure 1: Comparison between recovering a signal that belongs to the analysis model, with a frame as the analysis operator, using the standard sampling scheme with analysis algorithm and our new sampling scheme with synthesis algorithm. Left: Recovery rate for the noiseless case. Right: Reconstruction mean squared error in the noisy case.

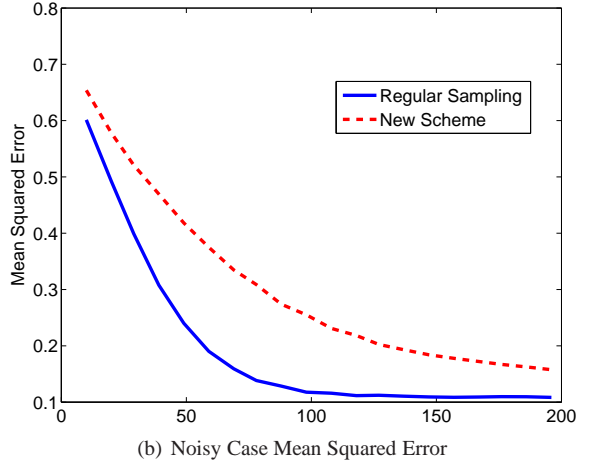
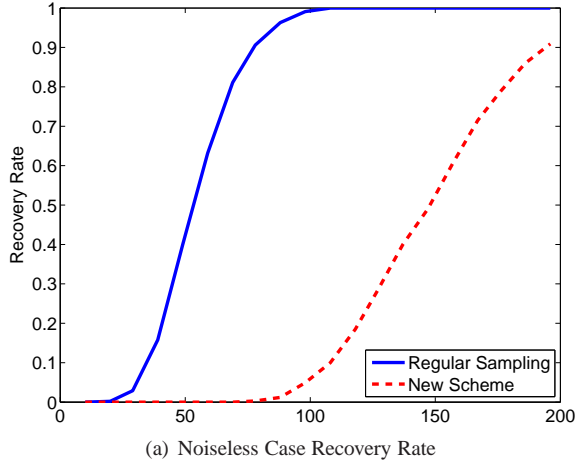


Figure 2: Comparison between recovering a signal that belongs to the analysis model, with the 2D-DIF as the analysis operator, using the standard sampling scheme with analysis algorithm and our new sampling scheme with synthesis algorithm. Left: Recovery rate for the noiseless case. Right: Reconstruction mean squared error in the noisy case.

be 10, see [32] for more details). In the standard sampling setup, the entries of the sensing matrix  $\mathbf{M} \in \mathbb{R}^{d\gamma \times d}$ , where  $\gamma \in \{0.05, 0.1, 0.15, \dots, 1\}$ , are randomly generated from an i.i.d random Gaussian distribution, followed by a normalization of each column to have a unit  $\ell_2$ -norm. For the new scheme we set  $\mathbf{M} = \mathbf{A}\mathbf{\Omega}$  with  $\mathbf{A} \in \mathbb{R}^{d\gamma \times 1.2d}$  a random Gaussian matrix selected in the same way that  $\mathbf{M}$  is selected in the standard sampling scheme. For each value of  $\gamma$  we generate 1000 different sensing matrices and signals  $\mathbf{x}$  that have sparsity  $k$  under  $\mathbf{\Omega}$ . The signals are generated by projecting a randomly selected Gaussian vector to the subspace orthogonal to randomly selected  $n - k$  rows from  $\mathbf{\Omega}$ , followed by normalization of the vector.

In Fig. 1 we present the recovery rate of the two algorithms in the noiseless and noisy cases. The noise is set to be i.i.d white Gaussian with  $\sigma = 0.01$ . It can be seen that it is possible to recover signals from the analysis model using Algorithm 1. However, this comes at the cost of using more samples in order to achieve the same recovery rate and error. This shows us that though the theoretical guarantees of the analysis algorithms take into account only  $k$  and not the intrinsic dimension of the signals, losing the information about the latter, which happens when we sample in

the transform domain, may harm the recovery. On the other hand, if we can afford having more measurements, then we have the privilege of using existing synthesis algorithms, which have a large variety of efficient implementations compared to what is available for the analysis model. For example, compare the methods available for the generic synthesis  $\ell_1$ -minimization problem [52, 53, 54, 55, 56, 57, 58, 59, 60, 61] to the ones designed for the generic analysis  $\ell_1$ -minimization [62, 63]. Remark that the advantage in efficiency is not unique to the  $\ell_1$ -relaxation alone. For more examples, we mention the sampling with expander graphs [44] that does not have a counterpart in the analysis framework and refer the reader to compare OMP with GAP [18] or the synthesis greedy-like algorithms with their analysis versions [32].

We repeat the experiment with the 2D-DIF operator and compare analysis  $\ell_1$ -minimization with the scheme in Algorithm 2 that uses synthesis  $\ell_1$ -minimization for recovery. The signals we generate are random  $14 \times 14$  images with four connected components. We start with a constant image and then add to it three additional connected components using a random walk on the image using the same technique in [19]. The sensing matrices are selected as in the previous experiment, where in the new sampling scheme we assign 2 measurements (from the total number of measurements we use) in the noiseless case for the signal recovery from the transform domain proxy and  $m/10$  in the noisy case.

Figure 2 presents the reconstruction rate in the noiseless case and the recovery error in the noisy case, where the noise is the same as in the previous experiments. We see the same phenomenon that we saw in the previous experiment but stronger. As the redundancy in the analysis operator is bigger in this experiment, the number of measurements we need for the new scheme is relatively larger and the recovery error in the noisy case is higher. Another reason, other than the bigger redundancy, for the inferior performance in this case is that we separate the measurements we have into two parts, where in the standard scheme the analysis  $\ell_1$ -minimization uses all the measurements at once for the recovery of the signal. Note that this causes that even in the case that  $m = d$  we do not get 100% recovery. Clearly in this case we will just invert the measurement matrix instead of using neither of the two schemes.

#### 4. Discussion and Conclusion

In this work we have presented a new sampling and recovery strategy for signals that are sparse under frames or the 2D-DIF operator in the analysis model. Our scheme utilizes existing algorithms from the synthesis sparsity model to recover signals that belong to the analysis framework. The advantage of this technique is that it enables the usage of existing tools for recovering signals from another model. Though in theory there is no additional cost for the usage of this scheme, it seems that in practice its advantage comes at the cost of the usage of more measurements in the sampling stage. This gap between the theory and practical performance gives us a hint that the existing guarantees are not tight and that there is a need for further investigation of the field. Another direction that should be further explored is the usage of the structure in the signals for designing the sampling operator, as is done for the 2D-DIF operator [27, 28].

#### Acknowledgment

The author would like to thank Michael Elad and Yaniv Plan for fruitful discussions. Raja Giryes is partially supported by AFOSR. The authors would like to thank the anonymous reviewers for their helpful and constructive comments that greatly contributed to improving this paper.

#### References

- [1] A. M. Bruckstein, D. L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM Review* 51 (1) (2009) 34–81.
- [2] E. Candès, Modern statistical estimation via oracle inequalities, *Acta Numerica* 15 (2006) 257–325.
- [3] G. Davis, S. Mallat, M. Avellaneda, Adaptive greedy approximations, *Journal of Constructive Approximation* 50 (1997) 57–98.
- [4] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 1st Edition, Springer Publishing Company, Incorporated, 2010.
- [5] H. R. Simon Foucart, *A Mathematical Introduction to Compressive Sensing*, 1st Edition, Springer Publishing Company, Incorporated, 2013.
- [6] S. Mallat, Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.* 41 (1993) 3397–3415.
- [7] E. Candès, T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory* 51 (12) (2005) 4203 – 4215.

- [8] D. L. Donoho, M. Elad, V. N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, *IEEE Trans. Inf. Theory* 52 (1) (2006) 6–18.
- [9] E. J. Candès, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies?, *IEEE Trans. Inf. Theory* 52 (12) (2006) 5406–5425.
- [10] D. Needell, J. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, *Appl. Comput. Harmon. Anal.* 26 (3) (2009) 301–321.
- [11] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, *IEEE Trans. Inf. Theory* 55 (5) (2009) 2230–2249.
- [12] T. Blumensath, M. Davies, Iterative hard thresholding for compressed sensing, *Appl. Comput. Harmon. Anal.* 27 (3) (2009) 265–274.
- [13] S. Foucart, Hard thresholding pursuit: an algorithm for compressive sensing, *SIAM J. Numer. Anal.* 49 (6) (2011) 2543–2563.
- [14] T. Zhang, Sparse recovery with orthogonal matching pursuit under RIP, *IEEE Trans. Inf. Theory* 57 (9) (2011) 6215–6221.
- [15] M. Rudelson, R. Vershynin, Sparse reconstruction by convex relaxation: Fourier and gaussian measurements, in: *Information Sciences and Systems, 2006 40th Annual Conference on*, 2006, pp. 207–212.
- [16] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, A simple proof of the restricted isometry property for random matrices, *Constructive Approximation* 28 (3) (2008) 253–263.
- [17] M. Elad, P. Milanfar, R. Rubinstein, Analysis versus synthesis in signal priors, *Inverse Problems* 23 (3) (2007) 947–968.
- [18] S. Nam, M. Davies, M. Elad, R. Gribonval, The cospase analysis model and algorithms, *Appl. Comput. Harmon. Anal.* 34 (1) (2013) 30–56.
- [19] R. Giryes, Y. Plan, R. Vershynin, On the effective measure of dimension in analysis cospase models, <http://arxiv.org/abs/1410.0989> (2014).
- [20] E. J. Candès, Y. C. Eldar, D. Needell, P. Randall, Compressed sensing with coherent and redundant dictionaries, *Appl. Comput. Harmon. Anal.* 31 (1) (2011) 59–73.
- [21] Y. Liu, T. Mi, S. Li, Compressed sensing with general frames via optimal-dual-based  $\ell_1$ -analysis, *IEEE Trans. Inf. Theory* 58 (7) (2012) 4201–4214.
- [22] R. Giryes, S. Nam, M. Elad, R. Gribonval, M. Davies, Greedy-like algorithms for the cospase analysis model, *Linear Algebra and its Applications* 441 (0) (2014) 22–60, special issue on sparse approximate solution of linear systems.
- [23] D. Needell, R. Ward, Stable image reconstruction using total variation minimization, *SIAM Journal on Imaging Sciences* 6 (2) (2013) 1035–1058.
- [24] D. Needell, R. Ward, Near-optimal compressed sensing guarantees for total variation minimization, *IEEE Trans. Img. Proc.* 22 (10) (2013) 3941–3949.
- [25] M. Kabanava, H. Rauhut, Analysis  $\ell_1$ -recovery with frames and Gaussian measurements, <http://arxiv.org/abs/1306.1356> (2014).
- [26] R. Giryes, A greedy algorithm for the analysis transform domain, to appear in *Neurocomputing*.
- [27] F. Krahmer, R. Ward, Stable and robust sampling strategies for compressive imaging, *IEEE Trans. Img. Proc.* 23 (2) (2014) 612–622.
- [28] C. Poon, On the role of total variation in compressed sensing, <http://arxiv.org/abs/1306.1356> (2014).
- [29] H. Rauhut, K. Schnass, P. Vandergheynst, Compressed sensing and redundant dictionaries, *IEEE Trans. Inf. Theory* 54 (5) (2008) 2210–2219.
- [30] D. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell^1$  minimization, *Proc. Nat. Aca. Sci.* 100 (5) (2003) 2197–2202.
- [31] M. Davenport, D. Needell, M. Wakin, Signal space CoSaMP for sparse recovery with redundant dictionaries, *IEEE Trans. Inf. Theory* 59 (10) (2013) 6820–6829.
- [32] R. Giryes, D. Needell, Greedy signal space methods for incoherence and beyond, *Appl. Comput. Harmon. Anal.* To appear.
- [33] R. Giryes, M. Elad, Can we allow linear dependencies in the dictionary in the synthesis framework?, in: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 5459–5463.
- [34] R. Giryes, M. Elad, Iterative hard thresholding for signal recovery using near optimal projections, in: *10th Int. Conf. on Sampling Theory Appl. (SAMPTA)*, 2013, pp. 212–215.
- [35] R. Giryes, M. Elad, OMP with highly coherent dictionaries, in: *10th Int. Conf. on Sampling Theory Appl. (SAMPTA)*, 2013, pp. 9–12.
- [36] C. Hegde, P. Indyk, L. Schmidt, Approximation-tolerant model-based compressive sensing, in: *ACM Symposium on Discrete Algorithms (SODA)*, 2014.
- [37] B. Ophir, M. Lustig, M. Elad, Multi-scale dictionary learning using wavelets, *IEEE Journal of Selected Topics in Signal Processing* 5 (5) (2011) 1014–1024.
- [38] S. Ravishanker, Y. Bresler, MR image reconstruction from highly undersampled k-space data by dictionary learning, *IEEE Trans. Medical Imaging* 30 (5) (2011) 1028–1041.
- [39] S. Ravishanker, Y. Bresler, Sparsifying transform learning for compressed sensing MRI, in: *IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, 2013, pp. 17–20.
- [40] E. Candès, T. Tao, The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ , *Annals Of Statistics* 35 (6) (2007) 2313–2351.
- [41] P. Bickel, Y. Ritov, A. Tsybakov, Simultaneous analysis of lasso and dantzig selector, *Annals of Statistics* 37 (4) (2009) 1705–1732.
- [42] Z. Ben-Haim, Y. Eldar, M. Elad, Coherence-based performance guarantees for estimating a sparse vector under random noise, *IEEE Trans. Signal Process.* 58 (10) (2010) 5030–5043.
- [43] R. Giryes, M. Elad, RIP-based near-oracle performance guarantees for SP, CoSaMP, and IHT, *IEEE Trans. Signal Process.* 60 (3) (2012) 1465–1468.
- [44] S. Jafarpour, W. Xu, B. Hassibi, R. Calderbank, Efficient and robust compressed sensing using optimized expander graphs, *IEEE Trans. Inf. Theory* 55 (9) (2009) 4299–4308.
- [45] M. Stojnic, F. Parvaresh, B. Hassibi, On the reconstruction of block-sparse signals with an optimal number of measurements, *IEEE Trans. Sig. Proc.* 57 (8) (2009) 3075–3085.
- [46] Y. C. Eldar, M. Mishali, Robust recovery of signals from a structured union of subspaces, *IEEE Trans. Inf. Theory* 55 (11) (2009) 5302–5316.
- [47] J. A. Tropp, A. C. Gilbert, M. J. Strauss, Algorithms for simultaneous sparse approximation. part i: Greedy pursuit, *Signal Process.* 86 (3) (2006) 572–588.
- [48] R. Baraniuk, V. Cevher, M. Duarte, C. Hegde, Model-based compressive sensing 56 (4) (2010) 1982–2001.

- [49] S. Hawe, M. Kleinstuber, K. Diepold, Analysis operator learning and its application to image reconstruction, *Image Processing, IEEE Transactions on* 22 (6) (2013) 2138–2150.
- [50] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1 (Mar. 2014).
- [51] M. Grant, S. Boyd, Graph implementations for nonsmooth convex programs, in: V. Blondel, S. Boyd, H. Kimura (Eds.), *Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences*, Springer-Verlag Limited, 2008, pp. 95–110.
- [52] M. Elad, B. Matalon, J. Shtok, M. Zibulevsky, A wide-angle view at iterated shrinkage algorithms, Vol. 6701, 2007, pp. 670102–670102–19.
- [53] E. T. Hale, W. Yin, Y. Zhang, Fixed-point continuation for  $\ell_1$ -minimization: Methodology and convergence, *SIAM Journal on Optimization* 19 (3) (2008) 1107–1130.
- [54] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for  $\ell_1$ -minimization with applications to compressed sensing, *SIAM J. Img. Sci.* 1 (1) (2008) 143–168.
- [55] S. Wright, R. Nowak, M. Figueiredo, Sparse reconstruction by separable approximation, *IEEE Trans. Sig. Proc.* 57 (7) (2009) 2479–2493.
- [56] E. van den Berg, M. P. Friedlander, Probing the pareto frontier for basis pursuit solutions, *SIAM Journal on Scientific Computing* 31 (2) (2009) 890–912.
- [57] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Journal on Imaging Sciences* 2 (1) (2009) 183–202.
- [58] J. Yang, Y. Zhang, Alternating direction algorithms for  $\ell_1$ -problems in compressive sensing, *SIAM Journal on Scientific Computing* 33 (1) (2011) 250–278.
- [59] S. Becker, J. Bobin, E. J. Cands, NESTA: A fast and accurate first-order method for sparse recovery, *SIAM Journal on Imaging Sciences* 4 (1) (2011) 1–39.
- [60] F. Bach, R. Jenatton, J. Mairal, G. Obozinski, Optimization with sparsity-inducing penalties, *Foundations and Trends in Machine Learning* 4 (1) (2011) 1–106.
- [61] E. Treister, I. Yavneh, A multilevel iterated-shrinkage approach to  $\ell_1$  penalized least-squares minimization, *Signal Processing, IEEE Transactions on* 60 (12) (2012) 6319–6329.
- [62] Z. Tan, Y. Eldar, A. Beck, A. Nehorai, Smoothing and decomposition for analysis sparse recovery, *IEEE. Trans. Sig. Proc.* 62 (7) (2014) 1762–1774.
- [63] R. Gu, A. Dogandzic, Reconstruction of nonnegative sparse signals using accelerated proximal-gradient algorithms, <http://arxiv.org/abs/1306.1356> (2014).